

INFORMATION RICH LIBRARIES

Inventors:

Volker Schellenberger
914 Moreno Avenue
Palo Alto, CA 94303
Citizenship: Germany

Donald P. Naki
1889 Sunset Boulevard
San Diego, CA 92103
Citizenship: U.S.A.

Thomas B. Morrison
3767 Redwood Circle
Palo Alto, CA 94306
Citizenship: U.S.A.

Prepared By:

McCutchen, Doyle, Brown, & Enersen, LLP
Three Embarcadero Center, Suite 1800
San Francisco, California 94111
(650) 849-4908

Express Mail Label No. EL893834681US

INFORMATION RICH LIBRARIES

5 CROSS-REFERENCES TO RELATED APPLICATIONS

This application claims the benefit of priority to U.S. Provisional Patent Application No. 60/239,476, filed October 10, 2000.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

Not Applicable.

TECHNICAL FIELD

This invention relates to methods for producing information rich polynucleotide libraries and articles and compositions useful therein and produced thereby.

BACKGROUND OF THE INVENTION

There is currently no effective way to systematically screen all possible permutations of a polymeric biological molecule such as a polynucleotide or protein for a property of interest where the molecule is of significant length. To test four nucleotides and 20 amino acids at each position in a polynucleotide or protein, respectively, rapidly leads to a geometric increase in the number of molecules to be tested such that available methods of synthesis, and even available volumes for testing, are quickly exceeded for even a small length of such a polymer. Furthermore, even if it were physically possible to screen all permutations of a sequence of a given length, the brute force nature of such an approach would result in a great deal of the effort expended being wasted in producing and characterizing molecules lacking the desired activity.

As a compromise, a number of different approaches have arisen to sample some of the diversity available in such polymeric biological molecules.

There are two well known methods for attempting to improve the function of a protein. In random mutagenesis, one introduces random mutations and then screens for mutants with a desirable change. Although introducing more mutations per gene increases the chances of finding genes with interesting functions, each mutation potentially leads to a non-functional

protein (for instance by interfering with folding). Thus, if in creating a protein variant library, one increases the average number of mutations per gene, one then also increases the fraction of genes in the library that encode proteins which lack function.

Another method utilizes recombination between homologous coding sequences. The key advantage of recombination over random mutagenesis is that it introduces mutations known to function in a homologous protein. As a result, one generates libraries which have a relatively large diversity yet still contain a large fraction of functional mutants. In other words, recombination uses the information contained in homologous sequences to introduce diversity into a protein of interest. However, diversity in recombination is limited by the kind of information it can utilize (*i.e.*, it uses only homologous sequences) and recombination is limited in the way it utilizes that information. For example, one has limited control over the selection of crossover points. In another example, recombination usually moves regions of a gene (10-1000 bp). It rarely moves an individual residue from one sequence into a homologous position in another sequence.

Systematic approaches to altering residues in biological polymers have been made. See, for example, the "SELEX" procedures described in Tuerk et al., *Proc Natl Acad Sci USA* 1992 Aug 1, 89(15):6988-92, and the screening for aptamers as described in Bock et al., *Nature* 1992 Feb 6, 355(6360):564-6. Pools of degenerate molecules are tested for a desired activity and the molecules possessing the greatest level of such activity can be propagated and subjected to further rounds of mutagenesis and selection. Again, however, it is not possible to test all permutations of a sequence of any significant length, so such techniques are limited by a type of "founder effect" controlled by the number of different molecules actually present in the starting population.

Systematic approaches to mutating every position in a protein have also been performed. However, the diversity at any given position is typically limited to a single change. Furthermore, such changes are typically made and assayed individually, are not made in the form of a library, and therefore do not test for multiple mutations which may be required for any given mutation to exhibit its potential activity. In some cases, a number of multiple mutants have been made at different positions throughout a protein. However, these are again typically predefined, and do not result in the production of a library of different polymers.